

# Hand-Gesture-based Human-Machine Interface System using Compressive Sensing

Tomás Mantecón, Ana Mantecón, Carlos R. del-Blanco, Fernando Jaureguizar and Narciso García

**Abstract**— A novel and robust vision-based human-machine interface system to naturally interact with computers/smart devices is proposed. The key contribution is the introduction of a Compressive Sensing technique to largely reduce the dimensionality of highly discriminative feature descriptors (computed from depth imagery), which originally have an excessive and inoperative high dimension to be applied to a Support Vector Machine based classifier. The experimental results prove the appropriateness of this approach for the proposed system.

**Keywords**— SVM, DLQP, LBP, Compressive Sensing, gesture recognition, human-machine interface.

## I. INTRODUCTION

Thanks to the advent of new and affordable depth sensors, a new generation of human-machine interfaces (HMI) is arising to easily and naturally control a wide range of consumer devices: personal computers, smart TVs, smart phones, and tablets. Regarding gesture-based HMI, depth imagery has two main advantages in comparison with color imagery: it is much more robust to illumination changes in the scene, and provides richer structural information of the hands.

Recently, Compressive Sensing (CS) theory has become very popular in the fields of hand gesture, face or human action recognition to obtain a real-time performance in devices of limited computational power, via the reduction of the dimensionality of the underlying recognition problem. Some state-of-the-art solutions use CS to generate a dictionary of gestures from a large dataset of gestures [1]. In other cases, they use CS in combination with a Bag of Feature (BoF) strategy [2]. There are works that have applied other dimensionally reduction techniques for vision-based recognition problems, such as Singular Value Decomposition (SVD) [3] and Principal Component Analysis (PCA) [4]. However, the achieved reduction is very dependent on the data structure, whereas CS is able to achieve impressive reduction rates independently on the data distribution.

In this paper, a gesture-based HMI system to control consumer electronic devices using depth imagery is proposed. The key contribution of the paper is the introduction of a CS stage to reduce the dimensionality of Depth Local Quantize Pattern (DLQP) feature descriptor [5], which are highly discriminative, but inoperative to use without any reduction

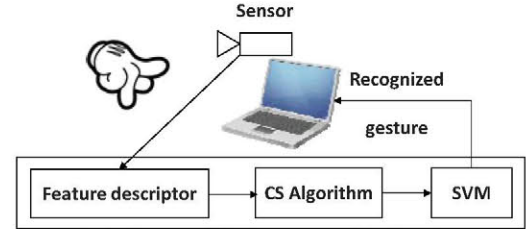


Fig. 1. Block diagram of the human-machine interface system.

because of their high dimensionality. Unlike other reduction techniques, CS is able to largely reduce the dimension of the descriptor while preserving the most of the information, since the relative distances among the reduced feature descriptors are (almost) preserved. The reduced DLQP is used as input in a bank of Support Vector Machine (SVM) classifiers that recognize the considered vocabulary of hand gestures to interact with an electronic device.

## II. SYSTEM DESCRIPTION

A block diagram of the proposed gesture-based HMI system is presented in Fig. 1. A depth sensor is used to acquire depth images of hand gestures. Then, a feature descriptor is computed for this images using the DLQP algorithm [5], which is inspired by the Local Binary Pattern (LBP) algorithm, but especially adapted to depth imagery via the inclusion of the quantized and coded magnitude of neighborhood differences, which are finally vectorized using a BoF approach. To reduce the high dimensionality of the obtained feature vector, the CS framework is adopted (see next section for more details). Finally, the reduced feature DLQP is delivered to a bank of SVM classifiers in one-vs-all configuration to perform the hand gesture recognition. To improve the performance of the classifiers, a Hellinger kernel, more commonly known as Bhattacharyya distance, is applied. Once a hand gesture is recognize, this triggers a command/action in the specific electronic device.

## III. COMPRESSIVE SENSING

Compressive Sensing (CS) [6] is a popular signal processing technique that allows to reconstruct sparse signals with fewer measurements than the ones required by the Nyquist-Shannon criterion. The main goal of applying this technique is to obtain a lower dimensionality vector with dimension  $M$  from a sparse vector with higher dimension

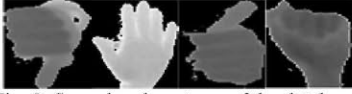


Fig. 2. Some hand gestures of the database.

$N \gg M$ . For this purpose, a measurement matrix  $\phi$  with dimensions  $(M \times N)$  is used to project the original vector  $x$  in a subspace, obtaining the reduced vector  $y = \phi \times x$ .

The matrix  $\phi$  must satisfy the Restricted Isometry Property (RIP) to preserve the distance among vectors, which in turns preserves the data structure, and therefore the underlying information. The RIP can be expressed as [7]:

$$(1 - \delta_k) \|x\|_2^2 \leq \|\phi x\|_2^2 \leq (1 + \delta_k) \|x\|_2^2 \quad (1)$$

where  $\|x\|_2$  is the Euclidean norm of  $x$ , and  $\delta_k \in (0,1)$  is the error in the vector distances after the projection  $\phi$ .

Different measurements matrices have been proposed in the literature to satisfy the RIP. One of those solutions is a random matrix whose elements are realizations of a Bernoulli distribution. The RIP is then satisfied if:

$$M \geq c \cdot K \cdot \log(N / K) \quad (2)$$

where  $c$  is a constant with a value close to 0.3, and  $K$  represents the number of elements of  $x$  that are not zero. Thus, a high sparsity,  $K$  small, allows higher reduction factors. This is the case of the DLQP descriptor, characterized by its huge sparsity.

The dimensionality reduction obtained by CS allows to reduce the computational cost and the memory requirements of the system, and also to perform a more accurate recognition thanks to the classification task is carried out in a lower dimensionality space.

#### IV. RESULTS

The depth-based hand gesture database called *depthGestRecog* [8] has been used to test the algorithm, which consists of a set of depth videos containing 11 different hand gestures, where each hand gesture has been performed by 6 different subjects.

A comparison using different configurations of the DLQP descriptor (that results in different feature lengths) has been performed. In addition, the whole system has been tested with and without the CS algorithm. The tested DLQP configurations are: NO\_CS (without CS), CS\_4 (with CS and 4 neighbors), CS\_8 (with CS and 8 neighbors), CS\_12 (with CS and 12 neighbors), and CS\_24, (with CS and 24 neighbors). TABLE I shows the length of the resulting feature vectors for several DLQP configurations and parameters of CS ( $M$ ,  $N$ , and  $K$ ).

TABLE I. VALUES OF  $M$ ,  $N$ ,  $K$  FOR DIFFERENT CONFIGURATIONS

	$N$	$K$	$M$
CS_4	65536	1483	1574
CS_8	131072	2662	2905
CS_12	196608	3891	4274
CS_24	393216	6790	7717

TABLE II. ACCURACY RESULTS USING DEPTHGESTRECOG DATABASE

Gesture	NO_CS	CS_4	CS_8	CS_12	CS_24
1	90,00	94,17	<b>100,00</b>	99,17	<b>100,00</b>
2	15,83	93,33	96,67	97,50	<b>98,33</b>
3	79,17	98,33	99,17	<b>100,00</b>	95,83
4	95,83	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	99,17
5	70,83	96,67	95,83	96,67	<b>98,33</b>
6	55,83	86,67	85,83	88,33	<b>92,50</b>
7	76,67	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>
8	60,83	95,00	96,67	95,00	<b>97,50</b>
9	<b>100,00</b>	99,17	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>
10	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>
11	38,33	<b>99,17</b>	93,33	<b>95,83</b>	90,83
Total	71,21	96,59	97,05	<b>97,50</b>	<b>97,50</b>

The adopted metric for the comparison is the accuracy:

$$\text{Accuracy} = 100 \times \frac{\text{Total number of correct hands}}{\text{Total number of hands}} \quad (3)$$

TABLE II shows the recognition accuracy for different configurations of the proposed system. The obtained results with the CS step experiment an accuracy increase of at least 25% with respect to the configuration that lacks of the CS step. Additionally, if we increase the size of the neighborhood from 4 pixels to 12, or 24 pixels, the accuracy improves up to 97%.

#### V. CONCLUSIONS

A robust gesture-based HMI system using depth imagery has been presented. A CS framework has been used to reduce the length of a DLQP descriptor, improving the computational cost, memory requirements, and recognition accuracy. The excellent obtained recognition scores using CS are attributed to the carefully design of the measurement matrix that creates a subspace where the vector distances are preserved.

#### REFERENCES

- [1] A. Boyali and M. Kavakli. "A robust gesture recognition algorithm based on sparse representation, random projections and compressed sensing". IEEE Conf. on Industrial Electronics and Applications, pp. 243-249, July 2012.
- [2] J.Wan, Q. Ruan, W. Li, and S. Deng. "One-shot learning gesture recognition from rgb-d data using bag of features". Journal of Machine Learning Research, vol. 14, pp. 2549-2582, September 2013.
- [3] R. Ptucha and A. Savakis. "Lge-ksvd: Flexible dictionary learning for optimized sparse representation classification". IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 854-861, June 2013.
- [4] S.-Z. Li, B. Yu, W. Wu, S.-Z. Su, and R.-R. Ji. "Feature learning based on sae-pca network for human gesture recognition in RGBD images". Neurocomputing, vol. 151, pp. 565-573, March 2015.
- [5] T. Mantecón, C.R. del Blanco, F. Jaureguizar, N. García, "Depth-Based Face Recognition using Local Quantized Patterns Adapted for Range Data", IEEE Int. Conf. on Image Processing pp. 293-297, October 2014.
- [6] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. "Introduction to compressed sensing". Preprint, April 2011.
- [7] D. Achlioptas. "Database-friendly random projections: Johnson-lindenstrauss with binary coins". Journal of Computer and System Sciences, vol. 66, pp. 671 - 687, June 2003.
- [8] <https://sites.google.com/site/depthgestrecog/>.